ACADEMIC PRESS

Available at WWW.MATHEMATICSWEB.ORG POWERED BY SCIENCE @ DIRECT®

Journal of Complexity 19 (2003) 548–554

Journal of COMPLEXITY

http://www.elsevier.com/locate/jco

Some remarks on the condition number of a real random square matrix

J.A. Cuesta-Albertos[a,1,*] and M. Wschebor[b,2]

[a] Departamento de Matemáticas, Estadística y Computación, Universidad de Cantabria, Spain

[b] Centro de Matemática, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay

Received 10 September 2002; accepted 29 January 2003

Abstract

In this paper we obtain some bounds for the expectation of the logarithm of the condition number of a random matrix whose elements are independent and identically distributed random variables. We also include some examples and extensions to cover the smoothed analysis as well as higher order moments.

MSC: 15A52; 65F35

Keywords: Condition number; Random matrices

1. Introduction

Let $A = (a_{i,j})_{i,j=1,\dots,n}$ be a real $n \times n$ nonsingular matrix. We denote by

$$||A|| = \sup_{||x||=1} ||Ax|| \tag{1}$$

the operator norm of A, where in the right-hand member of (1), "$||\cdot||$" is the euclidean norm in $\mathbb{R}^n$.

Define the *condition number of A* by

$$\mathscr{K}(A) = ||A|| \times ||A^{-1}||. \tag{2}$$

*Corresponding author.

E-mail addresses: cuestaj@unican.es (J.A. Cuesta-Albertos), wschebor@cmat.edu.uy (M. Wschebor).

[1] This author is partially supported by the Spanish Ministerio de Ciencia y Technología, Grants PB98-0369-C02-02 and BFM2002-04430-C02-02.

[2] Part of this work was supported by the Research Grants Council of the Hong Kong SAR (Project No. CityU 1085/02P).

0885-064X/03/$ - see front matter

doi:10.1016/S0885-064X(03)00010-4

The role of $\mathscr{K}(A)$ in numerical analysis—and specially in numerical linear algebra—has been recognized since a long time [6,7] (see also [3] and references therein).

In practice, it is useful to consider that the matrices we have to deal with are obtained in a random way (see [1,2,4,5]). We will assume that $\{a_{i,j}\}_{i,j=1,\dots,n}$ are independent and identically distributed (i.i.d.) random variables (r.v.'s) defined on the same probability space (Ω, σ, v) where the common distribution of $\{a_{i,j}\}_{i,j=1,\dots,n}$ depends on the kind of problem under consideration.

When matrix A is random, a certain number of natural questions about complexity of algorithms and effects of round-off errors leads to the study of the probability distribution of the random variable $\log \mathscr{K}(A)$ (note that $\mathscr{K}(A) \geqslant 1$ for any matrix A).

In what follows, we state and prove some elementary inequalities for the expectation and higher order moments of $\log \mathscr{K}(A)$, under quite general conditions on the randomness of A.

In the next section (Theorem 2.2) we prove the inequality for $E[\log \mathscr{K}(A)]$ which constitutes the main result of this paper. In Section 3 we present some applications and possible extensions of Theorem 2.2 including the bounds for higher order moments of $\log \mathscr{K}(A)$.

2. Main result

The proof employs the following simple and very well-known identity.

Lemma 2.1. *If X is a positive random variable defined on the probability space (Ω, σ, v), then*

$$E[X] = \int_0^\infty v[X > t]\, dt.$$

Theorem 2.2. *Assume that $a_{i,j}$, $i,j = 1, \dots, n$ are independent and identically distributed random variables and that their common probability distribution P satisfies the following conditions:*

1. *For any pair α, β of real numbers, $\alpha < \beta$, one has*
$$P([\alpha, \beta]) \leqslant P\left(\left[-\frac{\beta-\alpha}{2}, \frac{\beta-\alpha}{2}\right]\right). \tag{3}$$
2. $E[|a_{1,1}|^r] = \int_{-\infty}^{\infty} |x|^r P(dx) = 1$, *for some $r > 0$.*
3. *There exist positive numbers C, γ such that*
$$P([-\alpha, \alpha]) \leqslant C\alpha^\gamma \quad \textit{for all } \alpha > 0.$$

Then

$$E[\log \mathscr{K}(A)] \leqslant \left(1 + \frac{2}{r}\right) \log n + \frac{1}{r} + \frac{1}{\gamma}\{[(2+\gamma)\log n + \log C]^+ + 1\}, \tag{4}$$

where $x^+ = \max(x, 0)$ for real x.

Proof. Note that $||A|| \leqslant (\sum_{i,j=1}^n a_{i,j}^2)^{1/2}$. So, with the only assumption that the r.v.'s are i.d., for $t>0$

$$\begin{aligned} \nu[||A||>t] &\leqslant \nu\left[n^2 \sup_{i,j=1,\dots,n} a_{i,j}^2 > t^2\right] \\ &\leqslant \nu\left[\bigcup_{i,j=1}^n \left\{|a_{i,j}| > \frac{t}{n}\right\}\right] \leqslant n^2 \nu\left[|a_{1,1}| > \frac{t}{n}\right]. \end{aligned} \tag{5}$$

Hence, applying Lemma 2.1, for $\alpha_n \geqslant 0$

$$\begin{aligned} E[\log||A||] &\leqslant \alpha_n + \int_{\alpha_n}^{\infty} \nu[\log||A||>x]\, dx \\ &= \alpha_n + \int_{\alpha_n}^{\infty} \nu[||A||>e^x]\, dx \\ &\leqslant \alpha_n + \int_{\alpha_n}^{\infty} n^2 \nu\left[|a_{1,1}| > \frac{e^x}{n}\right] dx \\ &\leqslant \alpha_n + \int_{\alpha_n}^{\infty} n^2 \left(\frac{n}{e^x}\right)^r dx = \alpha_n + n^{2+r} \frac{1}{r} e^{-r\alpha_n}, \end{aligned} \tag{6}$$

where the last inequality follows from Markov's inequality and assumption 2.

Now choose $\alpha_n \geqslant 0$ to minimize the right-hand member of (6), i.e.

$$\alpha_n = \left(1 + \frac{2}{r}\right) \log n$$

and it follows that

$$E[\log||A||] \leqslant \left(1 + \frac{2}{r}\right) \log n + \frac{1}{r}. \tag{7}$$

Let us now consider the term $||A^{-1}||$. Denote $A^{-1} = (b_{i,j})_{i,j=1,\dots,n}$. Thus

$$b_{i,j} = \frac{a^{i,j}}{\det(\mathrm{A})}, \quad i,j = 1, \dots, n,$$

where $a^{i,j}$ is the adjoint of the position (i,j) in matrix A.

Clearly the r.v.'s $|b_{i,j}|, i,j = 1, \dots, n$ are i.d. and, so, we may apply (5) to matrix A^{-1} instead of A thus obtaining

$$\begin{aligned} \nu[||A^{-1}||>t] &\leqslant n^2 \nu\left[|b_{1,1}| > \frac{t}{n}\right] = n^2 \nu\left[\left|\frac{a^{1,1}}{\sum_{j=1}^n a_{1,j} a^{1,j}}\right| > \frac{t}{n}\right] \\ &= n^2 \nu\left[\left|a_{1,1} + \sum_{j=2}^n a_{1,j} \frac{a^{1,j}}{a^{1,1}}\right| < \frac{n}{t}\right]. \end{aligned}$$

The r.v.'s

$$a_{1,1} \quad \text{and} \quad \eta = \sum_{j=2}^{n} a_{1,j} \frac{a^{1,j}}{a^{1,1}}$$

are independent, so that, for each $\alpha > 0$, denoting by P_η the probability distribution of η, and using Fubini's theorem and assumption 1, we have

$$\begin{aligned} v[|a_{1,1} + \eta| < \alpha] &= \int_{-\infty}^{\infty} P[(-\alpha - y, \alpha - y)] P_\eta(dy) \\ &\leqslant \int_{-\infty}^{\infty} P[(-\alpha, \alpha)] P_\eta(dy) = P[(-\alpha, \alpha)]. \end{aligned}$$

Hence, by assumption 3,

$$v[||A^{-1}|| > t] \leqslant n^2 P\left(\left[-\frac{n}{t}, \frac{n}{t}\right]\right) \leqslant n^2 C\left(\frac{n}{t}\right)^{\gamma}, \tag{8}$$

and, with $\beta_n \geqslant 0$:

$$\begin{aligned} E[\log||A^{-1}||] &\leqslant \beta_n + \int_{\beta_n}^{\infty} v[||A^{-1}|| > e^x]\, dx \\ &\leqslant \beta_n + \int_{\beta_n}^{\infty} C n^{2+\gamma} e^{-\gamma x}\, dx = \beta_n + C\frac{n^{2+\gamma}}{\gamma} e^{-\gamma \beta_n}. \end{aligned}$$

Choosing

$$\beta_n = \frac{1}{\gamma}[(2+\gamma)\log n + \log C]^+,$$

one obtains

$$E[\log||A^{-1}||] \leqslant \frac{1}{\gamma}\{[(2+\gamma)\log n + \log C]^+ + 1\}, \tag{9}$$

and putting together (7) and (9), (4) is obtained. $\square$

Next we discuss briefly the assumptions in Theorem 2.2

Remark 2.2.1. It is not too difficult to show that assumption 1 implies that P is symmetric around 0. Therefore, if the r.v.'s $a_{i,j},\ i,j = 1, \dots, n$ are integrable, their common expectation must be 0.

Remark 2.2.2. Since $\mathscr{K}(\lambda A) = \mathscr{K}(A)$ for any real number λ and any nonsingular matrix A, in case $m_r = \int_{-\infty}^{\infty} |x|^r P(dx) < \infty$ it is possible to replace A by $m_r^{-1/r} A$ so that assumption 2 holds without modifying the condition number. Of course in this case one must change accordingly the constant C in assumption 3.

In this sense, assumption 2 is not more restrictive than the finiteness of the rth moment of the probability measure P.

3. Examples and extensions

3.1. Density function

Assume that P has a density function f, that f is even and nonincreasing on $[0, \infty)$ and that $m_r = \int_{-\infty}^{\infty} |x|^r f(x)\,dx < \infty$ for some $r > 0$.

We replace the original density f by $m_r^{1/r} f(m_r^{1/r} x)$ so that assumption 2 is satisfied without changing $\mathscr{K}(A)$; assumption 3 is verified with $\gamma = 1$ and $C = 2m_r^{1/r} f(0)$. Inequality (4) becomes

$$E[\log \mathscr{K}(A)] \leqslant \left(1 + \frac{2}{r}\right)\log n + \frac{1}{r} + \left[3\log n + \frac{1}{r}\log m_r + \log(2f(0))\right]^{+} + 1. \quad (10)$$

3.2. Uniform distribution

Consider the example in which P is the uniform distribution on $[-H, H], H > 0$. In this case, $m_r = H^r(r+1)^{-1}$ and (10) holds true for any $r > 0$. Letting $r \to +\infty$, we obtain

$$E[\log \mathscr{K}(A)] \leqslant 4\log n + 1. \quad (11)$$

3.3. Strong concentration near the mean

Here we analyze a family of distributions supported by $[-1, 1]$ but more concentrated around 0 as the uniform is. Assume that the density has the form

$$\frac{1}{2}\frac{\gamma}{|x|^{1-\gamma}}\mathbf{1}_{[-1,1]}(x),$$

for some γ, $0 < \gamma < 1$.

One has $m_r = \frac{\gamma}{r+\gamma}$ for each $r > 0$ and easily checks that introducing the modification suggested in Section 3.1 above, assumptions 1, 2 and 3 are satisfied with $C = m_r^{\gamma/r}$. Hence, Theorem 2.2 implies that for any $r > 0$

$$E[\log \mathscr{K}(A)] \leqslant \left(1 + \frac{2}{r}\right)\log n + \frac{1}{r} + \frac{1}{\gamma}\left\{\left[(2+\gamma)\log n + \frac{\gamma}{r}\log\frac{\gamma}{r+\gamma}\right]^{+} + 1\right\},$$

and, letting $r \to \infty$ it follows that

$$E[\log \mathscr{K}(A)] \leqslant \left(2 + \frac{2}{\gamma}\right)\log n + \frac{1}{\gamma}.$$

Notice that in this case, as $\gamma < 1$ the bound we obtain is worse than (11).

3.4. Gaussian distribution

The bound in Theorem 2.2 can be improved by using the actual distribution P instead of the Markov inequality in (6) or the bound in (8). This is, for example, the case for symmetric exponential or standard Gaussian distributions but we will not pursue the subject here. In the latter case, the precise behavior of $E[\log \mathscr{K}(A)]$ as $n \to \infty$ is given in [5] as

$$E[\log \mathscr{K}(A)] = \log n + C_0 + o(1),$$

where C_0 is a known constant.

3.5. "Smoothed analysis"

We now consider the condition number when the r.v.'s in the matrix $A = (a_{i,j})_{i,j=1,\dots,n}$ have the form

$$a_{i,j} = \mu_{i,j} + \psi_{i,j}, \quad i,j = 1, \dots, n,$$

where $M = (\mu_{i,j})_{i,j=1,\dots,n}$ is nonrandom and $(\psi_{i,j})_{i,j=1,\dots,n}$ are i.i.d. r.v.'s with common distribution P satisfying assumptions 1, 2 and 3 in Theorem 2.2. This—and other similar studies—have recently been called "smoothed analysis" (see [1,4]).

Theorem 3.1. *Under the conditions stated above, if*

$$\mu_n = \sup_{i,j=1,\dots,n} |\mu_{i,j}| \leqslant n^{2/r},$$

then

$$E[\log \mathscr{K}(A)] \leqslant \left(1 + \frac{2}{r}\right)\log n + \log 2 + \frac{1}{r} + \frac{1}{\gamma}\{[(2+\gamma)\log n + \log C]^+ + 1\}. \tag{12}$$

Proof. The proof (as well as the result) is very similar to that of Theorem 2.2. For $t > 0$ one has:

$$\begin{aligned} v[\|A\| > t] &\leqslant v\left[\sum_{i,j=1}^{n} a_{i,j}^2 > t^2\right] \leqslant \sum_{i,j=1}^{n} v\left[a_{i,j}^2 > \frac{t^2}{n^2}\right] \\ &= \sum_{i,j=1}^{n} v\left[|\mu_{i,j} + \psi_{i,j}| > \frac{t}{n}\right] \leqslant n^2 v\left[|\psi_{1,1}| > \frac{t}{n} - \mu_n\right]. \end{aligned}$$

Now choose $\alpha_n = (1 + \frac{2}{r})\log n + \log 2$ and check that if $x > \alpha_n$, then

$$\frac{e^x}{n} - \mu_n > \frac{1}{2n} e^x.$$

Thus,

$$\begin{aligned} E[\log\|A\|] &\leqslant \alpha_n + \int_{\alpha_n}^{\infty} v[\|A\| > e^x]\, dx \\ &\leqslant \alpha_n + n^2 \int_{\alpha_n}^{\infty} v\left[|\psi_{1,1}| > \frac{1}{2n} e^x\right] dx \end{aligned}$$

$$\leqslant \alpha_n + n^2 \int_{\alpha_n}^{\infty} \frac{1}{\left(\frac{1}{2n} e^x\right)^r} dx$$
$$= \left(1 + \frac{2}{r}\right) \log n + \log 2 + \frac{1}{r},$$

where last equality follows from a simple computation.

On the other hand, with the same notation as in the proof of Theorem 2.2, $A^{-1} = (b_{i,j})_{i,j=1,\dots,n}$ and

$$v\left[||A^{-1}|| > t\right] \leqslant \sum_{i,j=1}^{n} v\left[|b_{i,j}| > \frac{t}{n}\right].$$

For each term in this sum it is possible to repeat exactly the same computations as in the proof of Theorem 2.2 to bound $v\left[|b_{1,1}| > \frac{t}{n}\right]$ and obtain the same bound as there for $E[\log||A^{-1}||]$. This finishes the proof. $\square$

3.6. *Higher order moments*

It is possible to obtain upper bounds for $E[(\log \mathscr{K}(A))^k]$, $k = 2, 3, \dots$ much in the same way as we did for $k = 1$. We consider here the centered case, for smoothed analysis, the situation is similar.

Since $\log \mathscr{K}(A) \geqslant 0$ we have that

$$E[(\log \mathscr{K}(A))^k] \leqslant 2^k [E\{(\log^+ ||A||)^k\} + E\{(\log^+ ||A^{-1}||)^k\}].$$

Using the same estimates as in the case $k = 1$ for the tails of the probability distributions of $||A||$ and $||A^{-1}||$, after an elementary computation, it is possible to obtain that if $k \in \mathbb{N}$ satisfies that $2 \leqslant k \leqslant 1 + (2 + \gamma \wedge r) \log n$, then

$$E[(\log \mathscr{K}(A))^k] \leqslant (2 \log n)^k \left[\left(1 + \frac{2}{r}\right)^k (1 + k) + \left(1 + \frac{2}{\gamma}\right)^k (1 + Ck)\right].$$

References

[1] A. Blum, J. Dunagan, Smoothed analysis of the perceptron algorithm for linear programming, preprint 2002.
[2] D. Castro, J.L. Montaña, L.M. Pardo, J. San Martín, The distribution of condition numbers of rational data of bounded bit length, Found. Comput. Math. 2 (2002) 1–52.
[3] F. Cucker, Real computations with fake numbers, J. Complexity 18 (2002) 104–134.
[4] J. Dunagan, D.A. Spielman, S. Teng, Smoothed analysis of Renegar's condition number of linear programming, preprint, 2002.
[5] A. Edelman, Eigenvalues and condition numbers of random matrices, SIAM J. Matrix Anal. Appl. 9 (1988) 543–556.
[6] A. Turing, Rounding-off errors in matrix processes, Q. J. Mech. Appl. Math. 1 (1948) 287–308.
[7] J. von Neuman, H. Goldstine, Numerical inverting matrices of high order, Bull. Amer. Math. Soc. 53 (1947) 1021–1099.